Available online at www.sciencedirect.com
SCIENCE DIRECT
Remote Sensing of Environment 94 (2005) 123–132
Remote Sensing of Environment
www.elsevier.com/locate/rse

A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets

Chandra Giri[a,*], Zhiliang Zhu[b], Bradley Reed[a]

[a]SAIC/EROS Data Center, Sioux Falls, SD, 57198, USA
[b]U.S. Geological Survey, EROS Data Center Sioux Falls, SD, 57198, USA

Received 30 April 2004; received in revised form 5 September 2004; accepted 8 September 2004

Abstract

Accurate and up-to-date global land cover data sets are necessary for various global change research studies including climate change, biodiversity conservation, ecosystem assessment, and environmental modeling. In recent years, substantial advancement has been achieved in generating such data products. Yet, we are far from producing geospatially consistent high-quality data at an operational level. We compared the recently available Global Land Cover 2000 (GLC-2000) and MODerate resolution Imaging Spectrometer (MODIS) global land cover data to evaluate the similarities and differences in methodologies and results, and to identify areas of spatial agreement and disagreement. These two global land cover data sets were prepared using different data sources, classification systems, and methodologies, but using the same spatial resolution (i.e., 1 km) satellite data. Our analysis shows a general agreement at the class aggregate level except for savannas/shrublands, and wetlands. The disagreement, however, increases when comparing detailed land cover classes. Similarly, percent agreement between the two data sets was found to be highly variable among biomes. The identified areas of spatial agreement and disagreement will be useful for both data producers and users. Data producers may use the areas of spatial agreement for training area selection and pay special attention to areas of disagreement for further improvement in future land cover characterization and mapping. Users can conveniently use the findings in the areas of agreement, whereas users might need to verify the informaiton in the areas of disagreement with the help of secondary information. Learning from past experience and building on the existing infrastructure (e.g., regional networks), further research is necessary to (1) reduce ambiguity in land cover definitions, (2) increase availability of improved spatial, spectral, radiometric, and geometric resolution satellite data, and (3) develop advanced classification algorithms.

Keywords: Global land cover; Comparative analysis; Per-pixel comparison; Agreement and disagreement

1. Introduction

Two new global land cover data sets, Global Land Cover 2000 (GLC-2000) and MODerate resolution Imaging Spectrometer (MODIS) global land cover (MODIS land cover) have recently became available. The Joint Research Center (JRC) of the European Commission (EC) implemented the GLC-2000 project in partnership with more than 30 partner institutions around the world, using Satellite Pour l'Observation de la Terre (SPOT) VEGETATION 1-km satellite data (Fritz et al., 2003). Boston University prepared the MODIS land cover data using MODIS 1-km satellite data on board the Terra satellite (Friedl et al., 2002).

Both data sets were prepared with the same fundamental goal: to improve our understanding of the extent and distribution of the major land cover types of the world. The information generated can be used for various applications including ecosystem and biodiversity assessments, climate change studies, and environmental modeling (Brown, Loveland, Ohlen, & Zhu, 1999; Giri, Defourny, & Shrestha, 2003; Loveland & Belward, 1997; Loveland, Estes, & Scepan, 1999; Reed, 1997). The main objective of GLC-2000 was to prepare a harmonized land cover database

* Corresponding author. Tel.: +1 6055942835; fax: +1 6055946529.

0034-4257/$ - see front matter
doi:10.1016/j.rse.2004.09.005

of the world for the year 2000, primarily to serve international assessment programs such as the Millennium Ecosystem Assessment (MA) and the United Nation's Ecosystem-related International Conventions (GLC, 2003). The MODIS land cover was prepared for NASA's Earth Observing System (EOS) MODIS land science team.

These two global land cover data sets were prepared using different data sources, classification schemes, and methodologies, but using the same spatial resolution (1 km) satellite data. Similar classification efforts in the past show wide variations in the estimation of global land cover (Hansen & Reed, 2000; Townshend, Justice, Li, Gurney & Mcmanus, 1991). This is not surprising given the fact that quantitative analyses of complex land cover types remains an arduous task (Running, Loveland & Pierce, 1994; Zhu & Walter, 2003). Nevertheless, with the availability of improved spatial, spectral, geometric, and radiometric resolution satellite data (e.g., MODIS and VEGETATION), ground-truth data, and improved classification algorithms, it is possible to produce comprehensive and geospatially consistent global land cover data sets (Justice et al., 2002; Friedl et al., 2002).

The recent release of the GLC-2000 and MODIS land cover data sets call for a comparative analysis to examine their similarities and differences in terms of both methodology and results. It is also critical that both the data producers and users be aware of strengths and limitations of these data products. For example, the global area totals of land cover types could be similar; however, their spatial agreement could be vastly different. Furthermore, both area totals and spatial agreement could vary from region to region. From the producers' perspective, it is important to identify both areas of spatial agreement and disagreement. The areas of spatial agreement could be used as one of the ancillary data sets during training areas selection, and areas of disagreement could help identify issues for further improvement in future land cover characterization and mapping. In-depth understanding of similarities and differences will help users make informed decisions regarding the selection of global land cover data needed for their specific application. Similarly, users could conveniently utilize the data in the areas of agreement, whereas, in the areas of disagreement, users might need to verify the information with the help of secondary information.

The objective of this research is to summarize the similarities and differences in methodology and results of the GLC-2000 and MODIS land cover, and to identify areas of spatial agreement and disagreement.

2. Methodological similarities and differences

Before comparing the results, it is essential to understand the methodological similarities and differences between the GLC-2000 and MODIS land cover, which are summarized in Table 1.

Table 1
Characteristics of the GLC-2000 and MODIS land cover

Product characteristics	GLC-2000	MODIS land cover
Sensor	SPOT VEGETATION	Terra MODIS
Data used	Daily 1-km data acquired from 1 Nov. 1999 to 31 Dec. 2000 by SPOT-4 satellite, daily data consist of four spectral channels, and NDVI among others	Daily data acquired from 15 Oct. 2000 to 15 Oct. 2001, 250- and 500-m data resampled to 1 km for the first seven bands of MODIS data
Classification system	Flexible classification system depending on the partner institutions	Supervised classification system using decision tree classifier
Classification scheme	Flexible classification system using Land Cover Classification System developed by FAO and UNEP including both regional and global	IGBP, Global Ecosystem, UMD land cover, BGC biome scheme, LAI/fPAR biome scheme
Refinement/ update schedule	Currently in progress	Every 6 months
Results validation	Currently ongoing	Evaluated at global, continental, and individual class levels

The GLC-2000 was based primarily on SPOT VEGETATION daily 1-km data. However, some other data sources such as SPOT VEGETATION 10-day mosaic (S-10) data (of Southeast Asia), and the Normalized Difference Vegetation Index (NDVI), radar, and Defense Meteorological Satellite Program (DMSP) data (of Africa) were also used. The MODIS land cover data were prepared using MODIS surface reflectance (channels 1–7), MODIS Vegetation Index, MODIS Bidirectional Reflectance Distribution Function (BRDF), and other ancillary data. Two primary data sources, SPOT VEGETATION daily 1-km satellite data and MODIS data sets, were acquired in 1999/2000 and 2000/2001, respectively. We assume that the time difference of approximately 10 months between these two data sets is negligible in our comparative analysis.

The GLC-2000 used different classification techniques in different parts of the world depending on the requirements and preferences of partner institutions (Fritz et al., 2003); thus, the classification techniques are arguably optimized to regional and local needs (Mayaux, Bartholome, Massart, & Belward, 2002). In essence, each of the partner institutions was free to use any suitable and convenient classification technique. The classification techniques varied from supervised to unsupervised classification systems, digital to visual image processing techniques, or a combination of each. Fritz et al. (2003) provide a detailed description of the classification techniques used in each region/continent. In contrasts, MODIS land cover data used a consistent land cover classification system throughout the world. The project used a decision tree classifier with a supervised classification approach, using training samples "interpreted from high resolution imagery in association with ancillary data" (Friedl et al., 2002). The classification methodology of MODIS land cover is described in detail in Friedl et al. (2002) and Strahler et al. (1999).

The classification systems used also were different. The GLC-2000 used a flexible classification system based on the Land Cover Classification System (LCCS) developed by Food and Agriculture Organization (FAO) and the United Nations Environment Programme (UNEP). Regional participants were free to choose any number of land cover classes provided the minimum number of land cover classes (available at: URL http://www.gvm.sai.jrc.it/glc2000/Legend/GLC2000-LCCS_global-legend_overview.doc) required for global synthesis were included and the LCCS classification code was generated. In many cases, more detailed classifications were carried out at continental/regional levels than at the global level. Those land cover classes were combined at the continental/global level using LCCS code. A detailed description of the LCCS is described in Di Gregorio and Jansen (2000). The MODIS land cover primarily used the International Geosphere Biosphere Programme (IGBP) classification system (Loveland et al., 2000) along with other classification schemes such as the University of Maryland (UMD) land cover (Hansen & Reed, 2000), the BioGeoChemical (BGC) biome scheme (Running et al., 1994), and the Leaf Area Index/Fraction of Photosynthetically Active Radiation (LAI/fPAR) biome scheme (Myneni, Nemani, & Running, 1997). In addition, a prediction of the most likely alternative class of IGBP label also was generated.

The GLC-2000 was updated in various regions of the world based on comments received from users and data reviewers. Only minor updates are expected to be made in the future (S. Fritz, e-mail communication). The MODIS land cover data will be updated at quarterly (96-day) intervals (Strahler et al., 1999; Friedl et al., 2002). These quarterly updates are being performed to review and revise the existing maps by removing classification inconsistencies in earlier versions. Data in the future will be updated on a semi-annual or annual basis (Friedl et al., 2002).

The JRC and its partners around the world are presently validating the results of GLC-2000. Ideally, an independent expert not involved in the classification process will conduct the validation exercise (A. Belward, personal communication). The advantage of using a broad network of experts in validation is twofold: they are knowledgeable of their respective area or region and have access to a large collection of reference data available at the national/regional levels. The results validation adopts a two-step process: a confidence building method and a design-based method (GLC, 2003). In the confidence building method, a systematic review of the land cover product is performed by dividing the area into regularly spaced grids. The gross errors are identified and corrected using ancillary data such as thematic maps and satellite images through a series of breakout sessions during the results workshop. In the design-based approach, global land cover data will be compared with interpreted high-resolution satellite data, an approach similar to the validation of IGBP DISCover (Belward & Loveland, 1996). Visual interpretation will be performed to the high-resolution satellite imageries using LCCS classifiers. The sample sites are selected on the basis of stratified random sampling by land cover classes with a minimum number of observations required in each class. The MODIS land cover was validated using "unseen training sites" (Anonymous, 2004), and "confidence values aggregated by land cover class and continental region" (Anonymous, 2004). The accuracy of the IGBP layer is 75–80% at the global level, 70–85% at the continental level, and 60–95% at the individual class level.

3. Data sources and methodology

The GLC-2000 v1.1 data were downloaded from the worldwide web at URL http://www.gvm.sai.jrc.it/glc2000/ (last accessed 24 February 2004). We acquired the data in the Geographic Coordinate projection system, and reprojected it to Interrupted Goode Homolosine projection system. The MODIS land cover data were acquired from the Earth Observing System (EOS) Data Gateway, http://edcimswww.cr.usgs.gov/pub/imswelcome/ (last accessed 12 February 2004). The data, which are available in Hierarchical Data Format (HDF), were downloaded as tiles in Interrupted Goode Homolosine projection system. The individual tiles were mosaicked together for global coverage using MODIS Reprojection Tool 3.0 (available at URL http://lpdaac2.usgs.gov/landdaac/tools/modis/index.asp). Biome data were obtained from the World Wildlife Fund (WWF) in vector format. The data were converted into GRID and re-projected to Interrupted Goode Homolosine projection system.

We used the IGBP layer of the MODIS land cover which contains 17 land cover classes. In contrast, the GLC-2000 contains 22 land cover classes. The latter was aggregated into 17 classes to make it consistent with the MODIS land cover classification system. Table 2 shows the translation from GLC-2000 to IGBP land cover classes. The translation, however, is not straightforward. For example, the definitions of "forest" in the IGBP classification scheme and LCCS are different. IGBP defines forests as lands dominated by woody vegetation with a percent cover of 60% and higher, and height exceeding 2 m. In LCCS, the threshold for forest is >15% tree cover and tree height of greater than 3 m.

The IGBP land cover classes were generalized by aggregating the 17 classes into eight major land cover classes: forest, savannas/shrublands, grasslands, cropland, mosaic of cropland and natural vegetation, barren lands, urban, and wetlands. This was necessary to examine whether agreement is higher at the generalized class level

Table 2
IGBP land cover classes, GLC-2000 equivalent classes, and their class description

IGBP class	GLC-2000 equivalent	IGBP class description
Evergreen needleleaf forest	Tree cover, needle-leaved, evergreen	Lands dominated by wood vegetation with a percent cover >60% and height exceeding 2 m. Almost all trees remain green all year. Canopy is never without green foliage.
Evergreen broadleaf forest	Tree cover, broad-leaved, evergreen or tree cover, regularly flooded, fresh water and saline water.	Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 m. Almost all trees and shrubs remain green year round. Canopy is never without green foliage.
Deciduous needleleaf forest	Tree cover, needle-leaved, deciduous	Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 m. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
Deciduous broadleaf forest	Tree cover, broad-leaved, deciduous, closed	Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 m. Consists of broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
Mixed forest	Tree cover, mixed leaf type tree cover, burnt	Lands dominated by trees with a percent cover >60% and height exceeding 2 m. Consists of tree communities with interspersed mixtures or mosaics of the four forest types. None of the forest types exceeds 60% of landscape.
Closed shrubland	Mosaic: tree cover/other natural vegetation	Lands with woody vegetation less than 2 m tall and with shrub canopy cover >60%. The shrub foliage can be either evergreen or deciduous.
Open shrubland	Shrub cover, closed-open, evergreen	Lands with woody vegetation less than 2 m tall and with shrub canopy cover between 10% and 60%. The shrub foliage can be either evergreen or deciduous.
Woody savanna	Tree cover, broad-leaved, deciduous, open	Lands with herbaceous and other understory systems, and with forest canopy cover between 30% and 60%.The forest cover height exceeds 2 m.
Savanna	–	Lands with herbaceous and other understory systems, and with forest canopy cover between 10% and 30%. The forest cover height exceeds 2 m.
Grasslands	Herbaceous cover, closed-open	Lands with herbaceous types of cover. Tree and shrub cover is less than 10%.
Permanent wetlands	Regularly flooded shrub and/or herbaceous cover	Lands with a permanent mixture of water and herbaceous or woody vegetation. The vegetation can be present in either salt, brackish, or fresh water.
Cropland	Cultivated and managed areas	Lands covered with temporary crops followed by harvest and a bare soil period mosaic lands (e.g., single and multiple cropping systems). Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type.
Urban/built-up	Artificial surfaces and associated areas	Land covered by buildings and other man-made structures.
Cropland/natural vegetation mosaic	Cropland/tree cover/other natural vegetation	Lands with a mosaic of croplands, forests, shrubland, and grasslands in which no one component comprises more than 60% of the landscape.
Snow/ice	Snow and ice	Lands under snow/ice cover throughout the year.
Barren lands	Bare areas	Lands with exposed soil, sand, rocks, or snow and never has more than 10% vegetated cover during any time of the year.
Water	Water bodies (natural and artificial)	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt water bodies

than the IGBP class level. Global area totals were computed for aggregated land cover classes and also for the IGBP classes. Pixel-by-pixel comparisons also were performed between the generalized (with eight classes) GLC-2000 and MODIS land cover data. In doing so, pixels with the same land cover classes in both data sets retained their class values, whereas pixels with different land cover classes were labeled as areas of disagreement. The percent agreement and disagreement were calculated using the following equations:

$$\text{Overall percent agreement} = \sum_{i=1}^{r} \left(\frac{X_i}{Y_i}\right) \times 100 \qquad \text{(i)}$$

$$\text{Overall percent disagreement} = 1 - \sum_{i=1}^{r} \left(\frac{X_i}{Y_i}\right) \times 100 \qquad \text{(ii)}$$

Class percent agreement

$$= \left(\frac{\text{Area totals of class}_i \text{ in GLC} - 2000}{\text{Area totals of class}_i \text{ in MODIS}} \right) \times 100 \tag{iii}$$

where r=the number of classes; X_i=area total of class i in GLC-2000; Y_i=area total of class i in MODIS land cover.

4. Results and discussion

This section first compares the global area totals obtained from GLC-2000 and MODIS global land cover data for both generalized and IGBP land cover classes. Then a per-pixel comparison of aggregated classes and spatial comparisons by biome are discussed. Finally, strengths and weaknesses of both data sets are discussed.

4.1. Areal comparison

Fig. 1 shows the global area totals of eight aggregated classes, which were aggregated on the basis of major life forms. The global area totals are quite similar for all land cover classes except for shrublands/savannas and wetlands. The large percent difference for shrublands/savannas obtained in this study is similar to earlier findings reported by Hansen and Reed (2000). In this specific case, discrepancies can be explained by the difference in the definitions of "forest" in the IGBP classification scheme and LCCS. In IGBP classification scheme, forests are defined as lands dominated by woody vegetation with a percent cover of 60% and higher and height exceeding 2 m. In LCCS, the threshold for forest is >15% tree cover and tree height greater than 3 m. By definition, forests in GLC-2000 overlap with woody savannas/shrublands of IGBP land cover. Moreover, the absence of woody savannas in GLC-2000 also contributes to the difference. The closest class in GLC-2000 to woody savannas is tree cover, broad-leaved, deciduous, open. Wetlands represent approximately 1 percent of total land area of the world, thus making it difficult to delineate the boundaries with coarse resolution satellite data. The large percent difference also might have arisen due to the definition of wetlands itself.

The percent agreement of global area totals of forest, grasslands, croplands, urban lands, barren lands, and mosaic of croplands/natural vegetation are 91.07%, 81.56%, 86.5%, 93.33%, 96.60%, and 74.46%, respectively. The percent agreement for shrublands/savannas and wetlands are 57.90% and 36.66%, respectively.

As expected, the overall agreement decreases as we increase the number of classes for comparison and vice versa. The discrepancy is much higher when we compared the two data sets using IGBP classes as opposed to aggregated classes (Fig. 2). The major differences were found in mixed forest, closed shrubland, open shrublands, and permanent wetlands land cover classes. The global area total of open shrubland is higher in the MODIS land cover than in GLC-2000. In contrast, the global area totals of closed shrublands and permanent wetlands are higher in GLC-2000 than in MODIS land cover. The GLC-2000 has more forest areas in the category of evergreen needleleaf forest, deciduous needleleaf forest, and deciduous broadleaf forest than does the MODIS land cover. On the other hand, MODIS land cover depicts more forest areas in the categories of evergreen broadleaf forest and mixed forest than GLC-2000 does. Similarly, MODIS land cover has more woody savannas than GLC-2000 has. For savannas, grasslands, croplands, cropland/natural vegetation mosaic, urban and built-up, snow and ice, and barren land classes, GLC-2000 contains more area compared to MODIS land cover.

The difference of global area totals between GLC-2000 and MODIS land cover for aggregated forest is 8.93%. However, the difference is much higher when comparing detailed forest cover types. The differences in evergreen needleleaf forest, deciduous needleleaf forest, evergreen

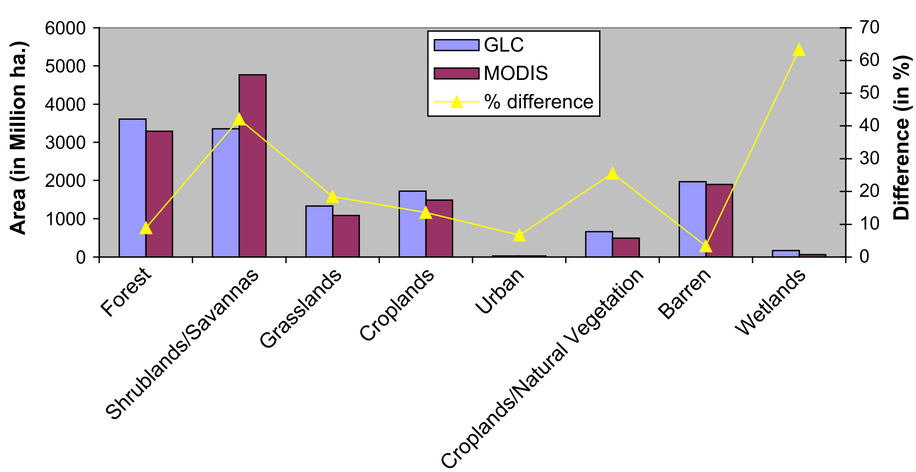

Fig. 1. Aggregated global area totals of GLC-2000 and MODIS land cover and their percent difference.

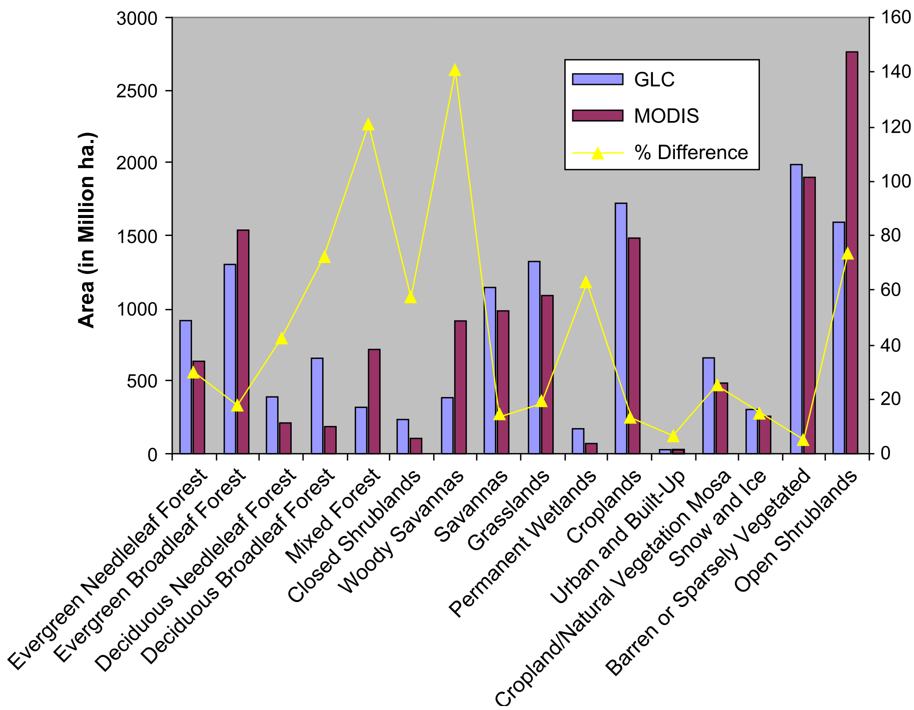

Fig. 2. Aggregated global area total for IGBP DISCover Classes.

broadleaf forest, deciduous broadleaf forest, and mixed forest are 30.00%, 17.63%, 42.76%, 72.31%, and 121.07%, respectively.

The differences in total areas for other detailed land cover types were variable: closed shrublands (57.33%), woody savannas (141.16%), savannas (13.95%), grasslands (18.44%), permanent wetlands (63.34%), croplands (13.50%), urban and built-up (6.67%), cropland/natural vegetation mosaic (25.54%), snow and ice (14.96%), barren or sparsely vegetated (3.40%), and open shrublands (73.76%).

4.2. Spatial (per-pixel) comparison

The per-pixel agreement between GLC-2000 and MODIS Land Cover is lower than global area totals. With aggregated land cover classes (eight land cover classes), the overall per-pixel agreement is 59.5% with a kappa coefficient of 0.48. Individual class agreement for forest, shrublands/savannas, cropland, and barrenlands are 69.4%, 67.6%, 52.1%, and 78%, respectively. In contrast, class agreement for grasslands, urban/built-up areas, croplands/natural vegetation, and wetlands are 21.8%, 34.5%, 9.6%, and 13.7%, respectively (Table 3).

Five major areas of disagreement between the two data sets occurs in (1) southern Siberia extending to the border of Kazakhstan, Mongolia, and China, (2) Sahel region of Africa, (3) southeastern part of Brazil, (4) Southern Australia, and (5) Tibetan plateau (Fig. 3). We have highlighted only the major disagreement areas; however, many other areas of disagreement do exist. Thus, users are cautioned to look at the differences in detail in their

Table 3
Confusion matrix showing agreement and disagreement between the GLC-2000 (row) and MODIS land cover (column)

GLC-2000/MODIS	Forest	Shrublands	Grasslands	Croplands	Urban	Cropland/natural vegetation	Barren	Wetlands	Row total
Forest	24 841 107	6 878 557	953 368	1 646 882	31 544	1 225 571	38 011	178 857	35 793 897
Shrublands	2 686 671	22 682 911	3 595 645	1 400 167	18 154	623 401	2 319 786	143 976	33 470 711
Grasslands	585 491	7 599 153	2 899 686	1 178 313	18 592	600 388	389 200	8918	13 279 741
Croplands	1 305 234	3 622 997	1 390 525	8 959 914	64 759	1 755 947	91 715	4096	17 195 187
Urban	26 676	35 324	10 398	83 665	96 428	19 206	7030	776	279 503
Cropland/Natural Vegetation	2 039 136	2 128 513	433 157	1 289 786	18 761	640 675	94 616	8335	6 652 979
Barren	23 442	2 917 996	1 332 928	59 063	1884	10 600	15 575 992	7081	19 928 986
Wetlands	562 511	626 007	118 528	104 412	3004	40 581	19 017	235 016	1 709 076
Column total	32 070 268	46 491 458	10 734 235	14 722 202	253 126	4 916 369	18 535 367	587 055	128 310 080

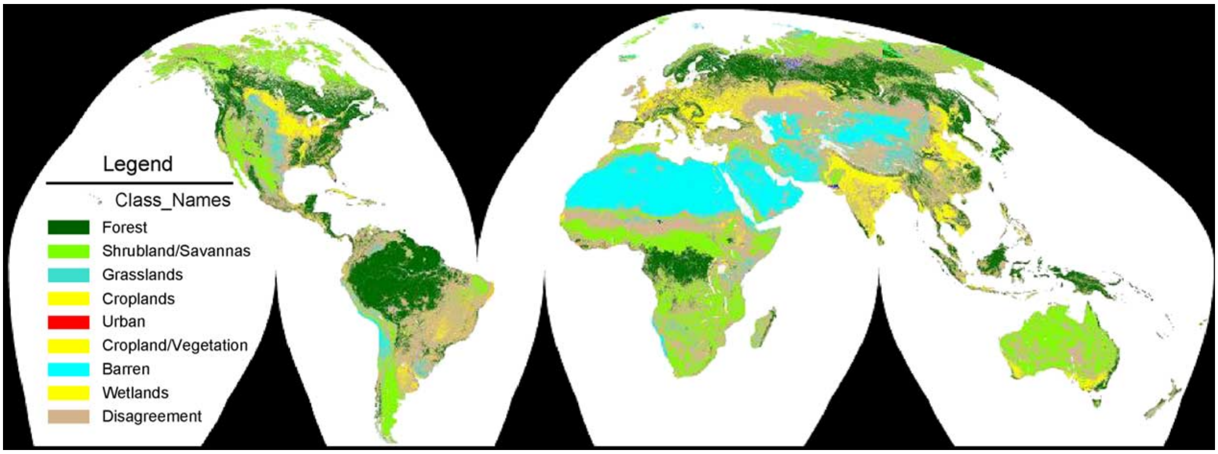

Fig. 3. Areas of agreement and disagreement in GLC-2000 and MODIS land cover.

particular area of interest. Areas of spatial agreement by land cover classes and areas of disagreement for the eight aggregated land cover classes are shown in Fig. 3.

In southern Siberia, MODIS land cover has grasslands, but GLC-2000 has grasslands in the southern belt and shrublands/woody savannas in the northern belt. This region is the world's largest steppe region encompassing grasslands, shrublands, and savanna belts stretching across central Asia from the Ural River in the west to the Altai foothills in the east. Because of the complex environment in southern Siberia, it is non-trivial to discriminate clear boundaries between grasslands, shrublands, and savannas. In the Sahel region of Africa, GLC-2000 shows regions of grasslands, woody savanna/shrublands, and croplands from north to south, whereas MODIS land cover shows woody savannas/shrublands with patches of grasslands in between. The Sahel region predominantly has sparse savanna vegetation of grasses and shrubs with very little precipitation. Sporadic patches of croplands also can be found depending on the annual precipitation.

Disagreement areas also can be found in the southeastern part of Brazil from the Atlantic Forest to the border of Brazil and Paraguay. More forest areas are in the MODIS land cover than in GLC-2000 for this region. The GLC-2000 depicts cropland and cropland/natural vegetation mosaic in much of the area. Similarly, in south Australia near Lake Eyre, more barren areas are in GLC-2000 than in MODIS land cover, where those areas are covered by woody savannas/shrublands in latter. Another area of disagreement is in western Australia where GLC-2000 shows woody savannas and grasslands, whereas MODIS land cover shows woody savannas/shrublands. The converse is true in the Tibetean Plateau, where MODIS land cover shows woody savannas/shrublands and grasslands, whereas the GLC-2000 shows all grasslands.

Disagreement also occurs in Alaska, U.S./Mexico border area, central part of Mexico, northern part of Columbia and Venezauela, Great Britain, central region of South Africa, India, Southeast Asia, and New Zealand.

Fig. 4 presents snapshots of areas of spatial agreement and disagreement between GLC-2000 and MODIS land cover (GLC-2000 is on the left and MODIS land cover is on the right). Fig. 4a represents an area in the vicinity of Sunderbans in the border area between Bangladesh and India. Here, there is a general agreement between mangrove forest and non-forest areas. Many patches of wetlands in the MODIS land cover may partly be because of overlapping definitions between mangrove forest and wetlands. Fig. 4b represents areas in and around the Olympic Mountains, Puget Sound, Seattle, and Tacoma of the United States. In this area, forest areas are well represented in both data sets but in the southeast corner of the image, the GLC-2000 shows shrublands/savanna while the MODIS land cover shows grasslands. Fig. 4c represents an area in southern part of South Africa. There are more forest areas in MODIS than in GLC-2000. Small patches of forest appear in GLC-2000 in the northeast part of the image, but are absent in MODIS land cover. This visual comparison strongly indicates the occurrence of regional variations between the GLC-2000 and MODIS land cover.

The discrepancies between the GLC-2000 and MODIS land cover can largely be explained with differing definitions used in the classification of SPOT VEGETATION and MODIS data. In addition, differences in input data sources, classification methodologies, and spatial details used in the image classification also are likely to have contributed to the overall discrepancy.

4.3. Comparison by biome

Areas of agreement and disagreement also were compared by biome to examine if high or low percent agreement is associated with any particular biome. This was necessary because some biomes are inherently complex in the extent and distribution of land cover

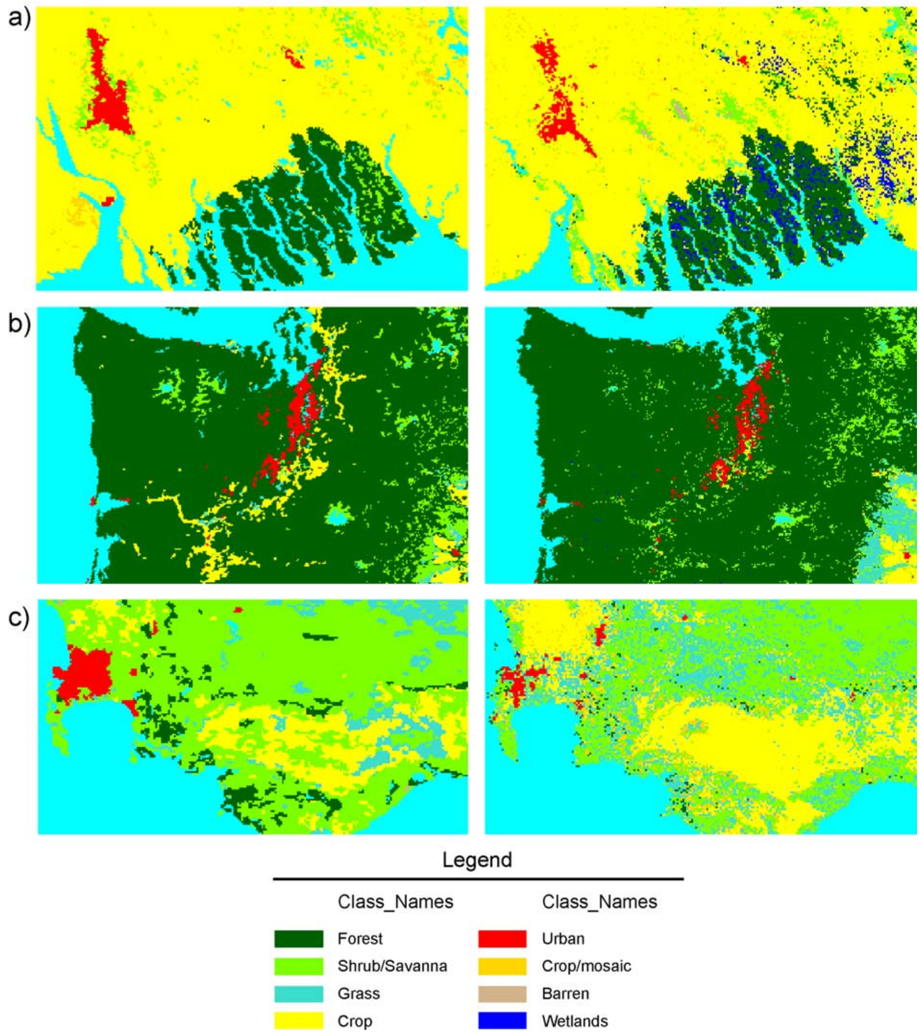

Fig. 4. Snapshots of areas depicting spatial agreement and disagreement between the GLC-2000 and MODIS global land cover data. The GLC-2000 is on the left and MODIS land cover is on the right. Selected area represents (a) an area in Sunderbans in the boarder of Bangladesh and India, (b) an area in and around the Olympic Mountains, Puget Sound, Seattle, and Tacoma of U.S., and (c) southern part of South Africa.

types, thus making them difficult to classify using coarse resolution satellite images.

The percent agreement between the two data sets was found to be highly variable among biomes (Table 4). A statistical test (Student's t-test) was performed to examine if percent agreement is consistently high or low for any particular biome. Our preliminary analyses suggest that the overall percent agreement is not significantly different by biome at 95% confidence interval. When performing the same test to the percent agreement by class, woody savannas/shrublands, and wetlands had significantly lower percent agreement compared to mean percent agreement of the total population. This analysis indicates that woody savannas/shrublands, and wetlands are difficult to discern with reasonable accuracy using either SPOT VEGETATION, or MODIS 1-km satellite data. A comparison of the IGBP DISCover and UMD land cover also revealed similar results (Hansen & Reed, 2000).

4.4. Major strengths and weaknesses

There are several advantages and disadvantages with the GLC-2000 and MODIS land cover characterization and mapping approaches. The major advantage of GLC-2000 is that the project was implemented with the active participation of more than 30 national, regional, and international organizations. Because of this overwhelming participation, it was possible to produce the GLC-2000 data at both regional and global scales. Participation of local and regional experts brought critical knowledge and experience of their respective regions while they participated in implementing GLC-2000. Active involvement of partner institutions also ensures international ownership of the data (Mayaux et al., 2002). In addition, the input data source, classification scheme, and classification methodology all were optimized to the needs of the participating institutions based on the land cover types found in their respective regions. The major weakness is that the methodology used

Table 4
Percent agreement by biome for eight major land cover types

Biome/land cover	Forest	Shrublands/savannas	Grasslands	Croplands	Urban	Croplands/natural vegetation	Barren	Wetlands
Tropical and subtropical moist broadleaf forests	53.29	59.49	34.78	45.50	46.63	27.99	46.48	25.92
Tropical and subtropical dry broadleaf forests	38.70	65.76	48.61	52.62	30.74	43.06	32.20	11.58
Tropical and subtropical coniferous forests	41.24	81.79	47.43	34.41	49.91	64.24	59.67	30.33
Temperate broadleaf and mixed forests	47.14	53.04	35.69	51.76	49.11	68.56	43.97	18.50
Temperate coniferous forests	46.07	60.78	59.36	48.11	48.86	74.86	34.59	23.88
Boreal forests/taiga	48.53	56.34	27.47	64.74	58.08	36.84	70.59	33.48
Tropical and subtropical grassland, savannas and shrublands	30.80	63.52	29.71	17.09	42.33	43.82	46.28	7.44
Temperate grasslands, savannas and shrublands	37.44	41.11	65.13	54.94	61.71	26.23	24.63	6.77
Flooded grasslands and savannas	38.30	66.30	42.54	49.75	53.05	41.86	52.76	18.58
Mountain grasslands and shrublands	33.02	69.59	35.75	21.59	22.68	37.59	55.89	13.43
Tundra	35.03	65.63	57.44	31.96	44.18	34.55	28.72	27.65
Mediterranean forests, woodlands and shrubs	26.97	53.04	54.05	46.09	35.37	81.86	59.16	39.75
Deserts and xeric shrublands	15.85	57.74	52.52	39.96	39.95	32.40	47.52	5.52
Mangrove	49.30	64.69	57.58	48.74	39.33	30.16	48.76	62.36
Lake	62.53	68.21	56.31	30.76	48.83	40.06	61.50	49.09
Rock and ice	42.74	54.73	17.63	19.25	91.67	28.64	70.06	25.00

to produce GLC-200 is not repeatable. Moreover, the level of effort varies in different regions depending on the time, resources, and expertise available to the participating institutions. Because of this, GLC-2000 varies in quality and/or level of details across regions/continents. The data are being validated using a standard accuracy assessment approach.

The major advantage of MODIS land cover data is that the project adopts a consistent methodology across the globe and is repeatable. The project aims to update the global data every 6 months by updating the training data and improving the classification algorithms. The major weakness of this approach is the lack of involvement of local/regional experts and lack of international ownership. The project does not intend to perform results validation using interpreted high-resolution satellite data citing lack of resources (Anonymous, 2004). The data were not intended for regional or local applications, thus the data were not optimized for those applications.

5. Conclusions

The foregoing comparative analyses provide insight for both data produces and users. For data producers, the identified areas of agreement may serve as a reference data for training areas selection. Likewise, areas of disagreement may receive special attention in future land cover characterization and mapping. Users also will have an opportunity to examine the similarities and differences in their area of interest, and make informed decisions based on their thematic applications. For example, users may conveniently use the data in the areas of agreement while they might have to verify the information in the areas of disagreement with the help of secondary information.

The purpose of this comparison is not to argue that one data set is better than the other, but to outline major similarities and differences highlighting their strengths and weaknesses. Clear understanding of similarities and differences in terms of global area totals for aggregated classes, IGBP classes, and also the spatial variability in different regions of the world is crucial before selecting data for particular applications.

Preliminary comparison of the GLC-2000 and MODIS land cover revealed that the overall global areal totals at class aggregate level are reasonably high except in the case of woody savannas/shrublands and wetlands. However, per-pixel agreement at class aggregate level is only 59.12%. This finding is not surprising given the fact that these data sets were prepared using different data sources, classification schemes, and methodologies. In fact, the result corroborates with earlier findings of similar comparisons conducted by DeFries and Townshend (1994), Hansen and Reed (2000), and Latifovic, Zhu, Cihlar, Giri, and Olthof (2003). The observed discrepancies might have arisen due to variable data sources, availability of ground-truth information, variable class definitions, and variable classification approaches used in the analyses. Also, errors might have been introduced due to the unavailability of cloud free images, mis-registrations, and other anomalies (Hansen & Reed, 2000).

Four time-series, 1-km spatial resolution global land cover data, namely 1992/1993 IGBP DISCover and UMD land cover and 2000/2001 GLC-2000 and MODIS land cover, are available for the first time. This encouraging development provides a solid foundation in generating geospatially consistent and accurate land cover characterization database of the world. Ironically, we will not be able to identify change areas comparing existing 1992/1993 and 2000/2001 data, simply because the "variability between estimates substantially exceeds that of actual land cover changes" (Defries & Townshend, 1994).

Learning from past experience and building on the existing infrastructure (e.g., regional network), the consistency and accuracy of global land cover data are expected to improve in the future. This will be facilitated by the availability of improved spatial, spectral, geometric, and radiometric resolution satellite data, superior classification methodologies, and better ground-truth information. Active involvement of local experts also can help produce an accurate and repeatable land cover characterization database. Further research is necessary to reduce ambiguity in land cover definitions, increase availability of improved spatial, spectral, radiometric, and geometric resolution satellite data, and to develop advance classification algorithms.

Acknowledgements

We would like to thank Craig Walters for editing the manuscript for language and clarity.

This work was made possible in part by Science Applications International Corporation under U.S. Geological Survey contract 1434-CR-97-CN-40274.

References

Anonymous. (2004). Validation of the consistent-year V003 MODIS land cover product. Available at: http://geography.bu.edu/landcover/userguidelc/consistent.htm (last accessed, 17 February, 2004).

Belward, A. S., & Loveland, T. R. (1996). The DIS 1 km land cover data set. *IGBP Global Change Newsletter*, *27*, 7–9.

Brown, J. F., Loveland, T. R., Ohlen, D. O., & Zhu, Z. (1999). The global land-cover characteristics database: The users' perspective. *Photogrammetric Engineering and Remote Sensing*, *65*(9), 1069–1074.

Defries, R. S., & Townshend, J. R. G. (1994). NDVI-derived land cover classification at global scales. *International Journal of Remote Sensing*, *15*, 3567–3586.

Di Gregorio, A., & Jansen, L. J. M. (2000). *Land cover classification system (LCCS)* (pp. 180). Rome: F.A.O.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, *83*(1–2), 287–302.

Fritz, S., Bartholome, E., Belward, A., Hartley, A., Stibig, H. -J., Eva, H., et al. (2003). *Harmonization, mosaicking, and production of the Global Land Cover 2000 database*. Ispra, Italy: Joint Research Center (JRC).

Giri, C., Defourny, P., & Shrestha, S. (2003). Land cover characterization and mapping of continental Southeast Asia using multi-resolution satellite sensor data. *International Journal of Remote Sensing*, *24*(21), 4181–4196.

GLC. (2003). Global Land Cover 2000 database. European Commission, Joint Research Centre. Available at: http://www.gvm.jrc.it/glc2000.

Hansen, M., & Reed, B. (2000). A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. *International Journal of Remote Sensing*, *21*(6–7), 1365–1373.

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., et al. (2002). An overview of MODIS land data processing and product status. *Remote Sensing of Environment*, *83*, 3–15.

Latifovic, R., Zhu, Z., Cihlar, J., Giri, C., & Olthof, I. (2003). Land cover mapping of North and Central America—Global Land Cover 2000. *Remote Sensing of Environment*, *89*, 116–127.

Loveland, T. R., & Belward, A. S. (1997). The international geosphere biosphere programme data and information system global land cover data set (DISCover). *Acta Astronautica*, *41*(4–10), 681–689.

Loveland, T. R., Estes, J. E., & Scepan, J. (1999). Global land cover mapping and validation—foreword. *Photogrammetric Engineering and Remote Sensing*, *65*(9), 1011–1012.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, J., Yang, L., et al. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR Data. *International Journal of Remote Sensing*, *21*(6/7), 1303–1330.

Mayaux, P., Bartholome, E., Massart, M., & Belward, A. S. (2002). The land cover of Africa for the year 2000. *LUCC Newsletter*, *8*.

Myneni, R. B., Nemani, R., & Running, S. W. (1997). Estimation of global leaf area index and absorbed PAR using radiative transfer models. *IEEE Transactions of Geoscience and Remote Sensing*, *35*, 1380–1393.

Reed, B. C. (1997). Applications of the U.S. Geological Survey's global land cover product. *Acta Astronautica*, *41*(4–10), 671–680.

Running, S. W., Loveland, T., & Pierce, L. L. (1994). A vegetation classification logic based on remote sensing for use in global biogeochemical models. *Ambio*, *23*, 77–81.

Strahler, A., Muchoney, D., Borak, J., Friedl, M., Gopal, S., Lambin, E., et al. (1999). MODIS land cover product: Algorithm Theoretical Basis Document (ATBD) Version 5.0. USA: Boston University.

Townshend, J., Justice, C., Li, W., Gurney, C., & McManus, J. (1991). Global land cover classification by remote sensing: present capabilities and future possibilities. *Remote Sensing of Environment*, *35*, 243–256.

Zhu, Z., & Walter, E. (2003). Global forest cover mapping for the United Nations food and agriculture organization forest resources assessment 2000 program. *Forest Science*, *49*(3), 369–380.